

Large Lineage-Specific Gene Expansions are Driven by Mobile Genetic Elements

Katherine H. Huang^{1,2*} (KHHuang@lbl.gov), Morgan N. Price^{1,2}, Eric J. Alm^{1,3}, Paramvir S. Dehal^{1,2}, Adam P. Arkin^{1,2,4}

¹Virtual Institute for Microbial Stress and Survival, <http://vimss.lbl.gov/>; ²Lawrence Berkeley National Laboratory, Berkeley, CA, 94720; ³Department of Biological Engineering, MIT, Cambridge, MA, 02139; and ⁴Department of Bioengineering, University of California, Berkeley, CA, 94720

Acknowledgements

This work was part of the Virtual Institute for Microbial Stress and Survival (<http://VIMSS.lbl.gov>) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics:GTL program through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy

Lineage-specific expansion (LSE) plays a vital role in how prokaryotes gain new gene functions and adapt to their environments. To uncover the mechanisms behind LSE, we identify genes that arise from LSE by constructing phylogenetic trees of protein families across 400+ bacterial genomes. We found that LSE genes tend to cluster on the chromosomes and form hyper LSE regions. Such regions could not be explained solely by operon duplication. The locations of these hyper LSE regions are often remarkably conserved among closely related strains, even though the gene content may not be conserved. Furthermore, these hyper LSE regions frequently overlap with clusters of mobile genetic elements (MGE) and strain-specific genomic islands. We hypothesize that the majority of large strain-specific gene duplications are mediated by MGE and are concentrated in regions prone to site-specific MGE-driven recombinations. And for the same reasons, these regions are more susceptible to phage integration and to intergenomic information exchange.